データドリブン手法を用いた空調熱源システムの流量および機器シミュレーションモデルの開発

金, 恵美
九州大学大学院人間環境学府空間システム専攻 : 博士後期課程

住吉, 大輔
九州大学大学院人間環境学研究院都市・建築学部門

https://doi.org/10.15017/4769758

出版情報：都市・建築学研究. 40, pp.19-34, 2021-07-15. 九州大学大学院人間環境学研究院都市・建築学部門

都市・建築学研究　九州大学大学院人間環境学研究院紀要　第40号，2021年７月
J. of Architecture and Urban Design, Kyushu University, No.40, pp.19～34, July. 2021

データドリブン手法を用いた空調熱源システムの流量および機器シミュレーションモデルの開発

Development of Flowrate and Equipment Simulation Model for Commercial Building HVAC&R System by Data-driven Method

金　恵美 *，住吉大輔 **
Hyemi KIM, Daisuke SUMIYOSHI

This study proposes a new simulation model of the HVAC&R system. The flowrate model directly calculates the flow value with the ANN model (black-box method), and the equipment performance model calculates the theoretical parameters with the ANN model and then performs the theoretical calculation (gray-box method). In addition, for the input variables of the ANN model, Pearson and Spearman correlation coefficients were used to select the input variables with the highest correlation for each model, and then calculations were performed. The accuracy of the Spearman correlation coefficient was high for both the flowrate model and the equipment model. Also, for the equipment model, the input variable based on the theoretical equation was as accurate as the input variable based on the Spearman correlation coefficient. Therefore, we found that systematic analysis of input variables affects the performance of the model.

Keywords: Data-driven Method, Gray-box Method, Machine Training, Simulation Model, Artificial Neural Network Model

1. Introduction

According to the presentation of the Global Alliance for Buildings and Construction[1], the building and construction sector accounted for 36% of the world's energy consumption as of 2019. Thus, the first effort to conserve energy in the building sector is to expand the introduction of energy management systems (EMS) for new and existing buildings. According to Grand View Research[2], the introduction of EMS in buildings is increasing as the size of the EMS market grows. This EMS is an indispensable system for building energy saving as it can grasp the detailed energy usage status of buildings in real-time and prevent reckless energy use and waste.

Another effort is to find ways to save energy using building energy models and simulation tools. In the construction field, energy is saved by using the energy performance simulation method, 1) introduction of new systems that use high-efficiency equipment and spend low energy consumption, 2) introduction of high-efficiency equipment that can save energy, and 3) effective operation methods for energy saving of existing buildings. However, two challenges must be addressed for these methods to be introduced and utilized more actively.

The first is that as the amount of data collected by EMS is vast and the scale of buildings and systems grows, it becomes difficult to understand the status of operations. To actively and properly utilize EMS, managers with a high level of expertise and data analysis capabilities for real-world buildings and systems must grasp real-time operation status and apply timely operation methods. However, there is a lack of such managers in the field, and it is difficult to find the optimal system operation in a building with only the judgment of a field expert.

The second task is that building energy performance simulation also requires a high level of expertise, and a lot of expert time and effort is required to obtain reliable simulation

*　空間システム専攻博士後期課程
**　都市・建築学部門

results. Operators can arbitrarily perform modeling on uncertain variables or insufficient input/output information during simulation based on individual subjective experience and judgment. So different modelers get different simulation results for the same building and system. Since this leads to reliability issues in simulation results, simulation tools and techniques need to be prepared based on more quantitative evidence.

This study proposes a new simulation model in which subjective intervention of modelers is minimized by using only the data collected when EMS is installed (hereinafter referred to as BEMS) in commercial buildings, and operation is predicted based on quantitative grounds. Since the model is calculated based on the collected data, inverse modeling that reproduces the current operation status of the central heat source system is possible, and based on the reproduced operation, it is possible to determine the correct operation and seek an optimized operation plan.

Currently, this study is developing a prototype technique, and the actual heat source system is very complex and diverse and has limitations such as loss of data and measurement errors collected from BEMS, which will have to addressed in the next research step. Therefore, a virtual heat source system was built using the existing simulation tool (e.g. TRNSYS), and the simulation result was assumed as actual measurement data, and then used for model development. The model can perform training using collected past data and predict the operation of the heat source system quickly. In addition, it aims to find an optimal driving plan that can save energy by using the prediction results of developed final entire simulation models.

In this paper, we establish detailed individual modeling techniques for the flowrate prediction model (hereinafter referred to as the flowrate model) and the equipment performance prediction model (hereinafter, the equipment model) using the data-driven method, and verify the performance of each model. The flowrate model predicts the water supply flowrate required for system operation for each equipment, and the equipment model predicts the operation of each equipment according to the load. We systematically established a data-based technique for the two models through a quantitative method using correlation coefficient analysis. These models can be applied to various central HVAC&R systems in the future.

2. Prior research on Gray-box Method in Building Energy Model

The data-driven method extracts the characteristics of collected data, grasping and predicting the configuration of a system and the correlation between variables. This method is divided into two types: a black-box and a gray-box. There is a method of selecting a model in full consideration of the characteristics of the data and optimizing the parameters inside the selected model to derive an answer with a small error. At this time, since it is difficult to perfectly track all of the parameters of the optimization process, the models are called the 'black-box method'.

However, in the case of a physical system, there is a problem with a calculation method that considers only the characteristics of the collected data. First, the collected data has a reliability problem. A high-performance, accurate sensor with very little error has been developed, but the actual measurement introducing such a sensor is still insufficient. Also, apart from the sensor's performance, errors by humans may occur in the process of installing the sensor and recording data. This leads to other problems, such as data loss and omissions.

Finally, in the case of a physical system, the relationship of the dependent variable according to the independent variable was established with several formulas. However, in the case of using the black-box method, the calculation of data is based on mathematical and statistical characteristics, not physical characteristics.

To compensate for this problem, an efficient method in which physical characteristics are combined with a black-box method is required when introducing it to a mechanical equipment system. Accordingly, a gray-box method was developed that combines the white-box method (the first law-based method) and the black-box method.

The Gray-box method indicates a method for simulating using both the physical theoretical formulas of systems and devices and machine learning. The gray-box method is subdivided into two. The first one is based on the premise of using a physical theoretical formula or an approximate formula and identifies some variables (parameters) contained in the formula using machine learning. The second method is to perform machine learning after narrowing down the selection of input variables based on physical relationships and professional judgment. In this case, the input variable that is closely related to the output variable is selected, but it is not guaranteed that the relationship between the input variable and the output variable follows the physical phenomenon. In this research, we aim to detect equipment faults in the transition of parameters identified by machine learning by combined the above two methods. It also has the advantage of being able to obtain output that conforms to physical phenomena.

In building simulation, there were rarely existed studies predicting the simple operation (e.g., operation signal, designed flowrate, etc.) of a heat source plant system by the black-box method. Therefore, we focus on previous studies using the gray-box method. Most of the studies directly predicted power consumption and load and they using the gray-box method had the characteristics of the above two methods. Therefore, this chapter focuses on the research on the existing gray-box method being studied in the field of building energy and simulation.

Pugliese[3] proposed a hybrid modeling approach that combines a gray-box model and a Gaussian process for interior modeling of office buildings. He modeled the building with TRNSYS and the HVAC system with MATLAB to predict the indoor temperature and energy consumption of the building. As a result, it has higher accuracy than the single modeling method.

Takagi[4] proposed a gray-box modeling method for the dynamic characteristics of building energy consumption for university buildings in Japan. Data collected during the heating period were used and parameters were estimated using nonlinear optimization. The secondary side heat capacity during heating was calculated using the ambient temperature and current of the refrigerator.

Braun[5] the field data to create an inverse model and used it to predict the heating and cooling loads on a building. He estimated the parameters for the secondary cooling load using a global direct search algorithm, and the optimal parameters were estimated using nonlinear optimization. He suggested that it is a modeling method with high accuracy.

Although the above studies have developed a gray-box technique for commercial buildings, studies on heating, ventilation, air-conditioning and refrigeration (HVAC&R) systems are insufficient. Previous studies focused on predicting the maximum indoor load in consideration of the thermal characteristics of the structure of the building, the number of occupants, and the amount of insolation. In other words, research on the gray-box method that directly targets the HVAC&R system of commercial buildings is still insufficient.

Others prior researches only developed individual modeling techniques. Lee[6], Estrada-Flores[7] and Hydeman[8] developed gray-box modeling for heat source equipment such as refrigerators and heat pumps, Arahal[9], Berkenkamp[10] and Tardioli[11] developed heat storage tanks, and Singh[12] developed cooling tower modeling. In the above studies, the modeling method is being developed centering on the equipment rather than the system operation, and thus, the research that applied the gray-box technique to the entire operation of the system targeting the entire HVAC&R system is insufficient.

Based on the above, this study targets the HVAC&R system of commercial buildings and develops a simulation modeling technique that combines a data-based method using collected data. This technique should be a modeling technique that can grasp the overall operation status of the HVAC&R system as well as the detailed operation status of each internal equipment. This can be said to seek an optimal operation plan from the macroscopic point of view (the whole system) and the microscopic point of view (individual equipment) to reduce the energy consumption of the HVAC&R system.

3. Overview of Target Systems and Data

This paper is a step to examine the applicability of the simulation model under development and uses virtual data without measurement errors before using the actual system data. However, the actual cooling load of the building measured from 2016 to 2017 in an office building that is the target of the simulation is used. The office building is a total floor area of 11,613.㎡ and five floors above ground, located in Takamatsu City, Kagawa Prefecture. The building is equipped with BEMS, but it is impossible to review the central heat source system, which is the subject of this study, as new and renewable energy and special systems are introduced. Accordingly, an air heat source type central heat source system capable of handling the maximum cooling load (measure value) of the target building was created using TRNSYS, and virtual operation data (hereinafter, operation data) was collected.

3.1 System Overview

The target system was only for the cooling period operation of the central heat source system using an air heat source. In addition, to reduce the peak load during the cooling period, a contraction heat system was introduced that stores the cooling heat source at night and dissipates it during the day. Two air-cooled chillers are introduced, one is operated to generate the heat source of the shrinking heat tank, and the other is operated

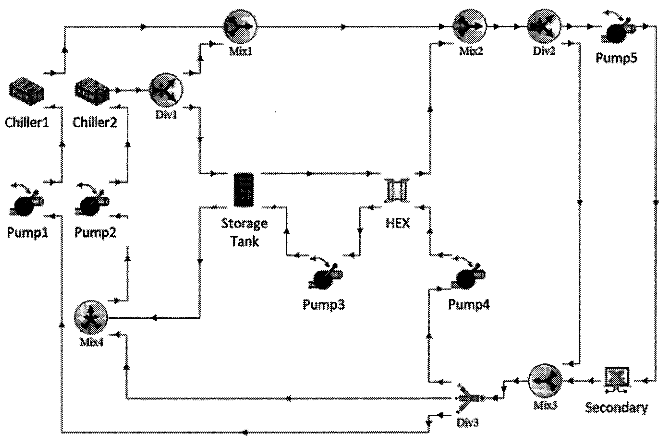

Fig. 1 Schematic of TRNSYS HVAC&R System

in stages to handle additional loads when the heat dissipation amount of the shrinking heat tank is exhausted. The detailed system diagram is shown in Fig. 1. There are five variable flow pumps for sending and receiving chilled water are installed for each piece of equipment.

3.2 Operation Data Overview

To collect data on the operation of the heat source system, in this study, virtual HVAC system operation data was created by substituting the cooling load generated in the actual building. The collected air-conditioning load data was used from July to September 2016 and from July to September 2017 by six months. The measurement interval of the data is one minute, and that of the virtual data created by TRNSYS is too.

The instantaneous maximum cooling load of the training data (2016) is 1039.55kW, and the predicted data (2017) is 827.78kW. In the case of outside temperature, the training data was distributed within 19.6~36.65℃, and the predicted data was distributed within 15.4~36.85℃. In the case of the relative humidity of the outside air, the training data was 24~98%, and the predicted data was 23~96%. The characteristics of the data in the prediction period can be said to have a value within the range of the data in the training period. However, in the case of the outside temperature, the maximum value was found to be as high as 0.2°C, but this was judged to be a sufficiently predictable range.

The virtual data that simulated the heat source system with TRNSYS using the actual cooling load data for three months in 2016 was used as the training data, and the data for the first one week were deleted in consideration of the stabilization of the TRNSYS calculation. There was a total of 122,400 training data, and 20% of the data were used as validation data. The operation data of the virtual heat source system for three months in 2017 was regarded as the data of the test, that is, the prediction period.

4. Overview of The Developed Simulation Model

4.1 Composition of The Entire Model

This study aims to minimize the modeler's subjective intervention and work time using only collected data and develop a simulation model of a heat source system with high accuracy. The model is largely composed of a flowrate model, an equipment operation model, and an equipment-to-equipment connection model, a total of three sub-models (Fig. 2). All three models are trained using historical data, and then the entire system operation is predicted in one-minute increments for the prediction period.

The flowrate model (Fig. 2 (a)) is a model for predicting the amount of water required for each operation of each piece of equipment by using variables that affect the secondary side building load and operation control (for example, heat source equipment temperature). Since the load of the air conditioning system is handled according to the flowrate and temperature of the chilled water supplied from each equipment of the heat source system, predicting the flowrate of each equipment can be said to be a measure that determines the detailed operation of each equipment. Hence, the flowrate model is used to calculate the amount of water supply during partial load operation of the heat source equipment and the amount of heat storage required during a heat storage operation.

Next, the equipment operation model (Fig. 2 (b)) is a model that predicts the outlet temperature and performance by using the flow rate and inlet temperature required to handle the load. The developed equipment modeling technique reflects the physical theories of each equipment because can be extended and used for future studies such as fault detection and diagnosis. In addition, it can be said to be a non-calibrated model that does not undergo a separate calibration process because it enables more detailed parameter calculation than an equipment model using the existing gray-box method.

Finally, there is a connection model between heat source equipment, storage tank and heat exchanger, etc. (Fig. 2 (c)). The connection model predicts the operation control of the heat source system and grasps the operation control according to the time, outdoor conditions, and past operation status. This model learns the relationship between the inlet and outlet temperatures between heat source equipment, storage tank and heat exchanger, etc. to determine and operate data transport during

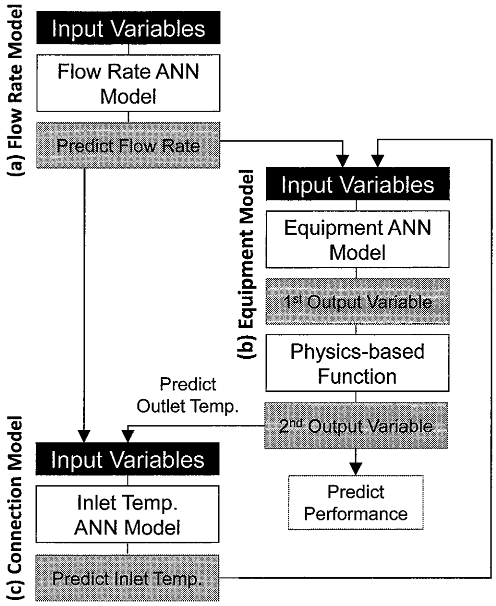

Fig. 2 Construction of the entire simulation model

prediction calculations. For example, during a heat storage operation, the temperature of chilled water discharged from the heat storage equipment is directed to the inlet on the low-temperature side of the heat storage tank, and the connection model automatically grasps and calculates this connection relationship. In this way, the connection model can be avoided from the hassle of manually inputting the operation schedule of the existing simulation method and grasping and modeling equipment components based on the system diagram.

Based on the characteristics of the above sub-models, the automated operation prediction model of the heat source system under development has the following characteristics.

(1) The entire calculation process is carried out automatically based on the characteristics of the data: automation

(2) The flowrate, temperature, equipment performance, and operation control inside the heat source system can be individually and specifically predicted: precision and high performance

(3) In the case of the equipment model, the correction is minimized by calculating the instantaneous parameters: non-calibration model

A developed simulation model with such characteristics means that it can automatically grasp (analyze) and predict the operation of the system from vast amounts of data. It can be easily used by non-professionals, and since it relieves the labor and burden of the operator, high use effects can be expected if enough accuracy is established. In this paper, detailed modeling methods for the flowrate model and equipment model were described and verified.

4.2 Sub-model Modeling Technique

The entire model is calculated based on the data-based modeling technique. As mentioned in Chapter 2, data-based modeling techniques include a black-box method based on interrelationships between data and a gray-box (or hybrid) model reflecting the physical characteristics of the system. In this study, two methods were customized to suit the characteristics of the sub-model. In addition, it aims to be able to cope with various central HVAC&R system by systematically establishing the modeling process.

First, the flowrate model can grasp the rough operation of the HVAC&R system using only the collected data, and the detailed operation of the system can grasp from the derived flowrate. However, the dataset used in this study is only temperature and flowrate. Therefore, there is no physical correlation to calculate the flow rate for each piece of equipment. For this reason, adopted the black-box model method for the flow rate model. This model is built by understanding how various independent variables affect specific values (dependent variables) across the entire range of collected data, rather than physically correlated.

For the equipment model, parameters (parameters that change according to external conditions and affect the performance) are derived based on the performance calculation theory of the equipment, and then the machine learning model calculates the parameters suitable for the theoretical expression for each data interval. At this time, the correlation between variables can be traced, enabling equipment fault detection and diagnostics (FDD). Therefore, the device model adopted the gray box model to predict the theoretical parameters.

The Artificial Neural Network (ANN) model was commonly used as a calculation technique for the black and gray-box methods of the two sub-models. The applied ANN model introduced the error backpropagation function based on the multiple layer perceptron (MLP) model with multiple hidden layers, and the error was optimized with the adaptive moment estimation (Adam) algorithm[13]. Three hidden layers and the number of nodes were set at 200, 100, and 50 respectively. For all nodes, the activation function is rectified linear unit (ReLU)[14]. By dividing the verification data by 20% by K-fold cross-validation, the verification process was performed five times to select an optimized model, and finally, after training the entire data, the prediction was performed.

4.3 Flowrate Model

For each of the five variable flow pumps in the target system, the estimated flowrate required to handle the load is calculated. Since the independent variables affecting individual pumps are different, which means that the weights of the ANN model should be applied differently. Hence, the single output model was applied without calculating the predicted flow values of five pumps at once and calculated each flowrate. Even when applying data from another new HVAC&R system, a single output ANN model will be created for each number of pumps to be calculated and will be computed every single ANN model in parallel.

Before creating individual ANN models for each pump, we selected input variables (independent variables). This is to quantitatively confirm that the variables affecting each flowrate of the pump are different, reduce model errors, and organize the modeling process. The input variables that can be used in predicting the flowrate were finally calculated through each input variable selection method in Section 5.1. Using the selected input variables, training and verification are performed with the ANN model. After that, each flowrate was predicted and verified using the optimized ANN model.

4.4 Equipment Performance Model

As mentioned in Chapter 2, there were two approaches to the gray-box method. The gray-box method used in this study combines the characteristics of two types. It reproduces the target device by predicting some parameters by machine learning based on the physical theoretical formula. Inputs to be used for machine learning We are improving the accuracy of machine learning by automatically selecting variable selections to be closely related to the output.

Air-cooled Chiller, heat storage tanks, and variable flow pumps in the target system of this study were modeled. After deriving the theoretical parameters for each piece of equipment, a dynamic calculation was performed using a data-driven technique. The nomenclature and calculation method for calculating the theoretical parameters for each piece of equipment is as follows, and the nomenclature is written down on the last page.

4.4.1 Air-Cooled Turbo Chiller

A TRNSYS Type 118 Air-Cooled turbo chiller was adopted as the heat source in this study. First, the chilled water outlet temperature parameter (hereinafter, $Pmt_{T_{evap}^{out}}$) for calculating the chilled water outlet temperature is determined using Eq. (1). The equation was studied by Gordon[15] and is called the simple thermodynamic model.

$$\frac{T_{evap}^{in}}{T_{cond}^{in}}\left(1+\frac{1}{COP}\right)-1=\frac{T_{evap}^{in}}{Q_{evap}}\Delta S_T+Q_{leak,eqv}\frac{\left(T_{cond}^{in}-T_{evap}^{in}\right)}{T_{cond}^{in}\times Q_{evap}}+\frac{R\times Q_{evap}}{T_{cond}^{in}}\left(1+\frac{1}{COP}\right) \quad \text{Eq. (1)}$$

$$Pmt_{T_{evap}^{out}}=T_{evap}^{in}\cdot P_R-\Delta S_T\cdot T_{evap}^{in}\cdot T_{cond}^{in}-Q_{leak,eqv}\left(T_{evap}^{in}-T_{cond}^{in}\right) \quad \text{Eq. (2)}$$

Furthermore, using the catalog value[16] in TRNSYS, the relationship between the cooling capacities, the inlet temperature on the heat source side, and the COP is generated by Eq. (3), which is a multiple regression equation.

$$COP_{TRNCata}=\beta_0+\beta_1 T_{cond,TRNCata}^{in}+\beta_2 Q_{evap,TRNCata}+\beta_{11}{T_{cond,TRNCata}^{in}}^2+\beta_{22}{Q_{evap,TRNCata}}^2+\beta_{12}T_{cond,TRNCata}^{in}Q_{evap,TRNCata}+\varepsilon \quad \text{Eq. (3)}$$

$$\mathrm{COP}_{TRNSYS}=COP_{TRNCata}\times Pmt_{COP} \quad \text{Eq. (4)}$$

The discrepancy between the catalog values of the COP generated using Eq. (3) and the COP during the actual operation (i.e., COP of TRNSYS result) is set as the COP parameter (hereinafter, Pmt_{COP}, according to Eq. (4)). In addition, $Pmt_{T_{evap}^{out}}$ and Pmt_{COP} are trained using the MLP model, and the two parameters are then predicted using the data on the prediction period. By substituting the predicted parameters into the theoretical formula for calculating the performance of the heat source equipment, the chilled water outlet temperature, COP, and power are calculated.

4.4.2 Heat Storage Tank

The heat storage tank is calculated using Eq. (6). The outlet temperature on the heat source equipment and load side is assumed to be the inlet temperature of the heat storage tank. Further, the temperature of the maximum temperature tank or the minimum temperature tank becomes the inlet temperature to the heat source equipment or the load side depending on the flow direction. The value of the heat loss in the tank is used as a theoretical parameter.

$$mC_W\frac{dT_{all}}{dt}=\dot{m}_sC_W(Tank_{SIT}-T_1)+\sum_{i=1}^{n}(\dot{m}_eC_W(T_{i-1}-T_i)-UA_s\sum_{i=1}^{n}(T_i-n\cdot T_a)+\frac{A_q\lambda_w}{z}\sum_{i=1}^{n}[(T_{i-1}-T_i)-(T_i-T_{i+1})] \quad \text{Eq.(6)}$$

4.4.3 Variable Flowrate Pump

The types of pumps are largely divided into a constant flow and variable flow pumps. Type 741 was adopted as the variable flow pump. The theory for calculating the shaft power of the pump is given by Eq. (6). The efficiency and head of the motor act as parameters in the following equation:

$$Power_{pump}=\frac{\rho QH}{1.02\eta} \quad \text{Eq. (7)}$$

$$Pmt_{Pump}=\frac{H}{\eta} \quad \text{Eq. (8)}$$

5. Selection of Input Variables

Since the characteristics of the flowrate model and the equipment model are different, and each model has not been sufficiently reviewed in a previous study, the importance and value of input variable selection are quantitatively analyzed to determine their importance and value. When applying machine training methods to other fields, Snieder[17] considers the characteristics of the data and thoroughly examines the relationships between them to create a machine training model. However, in the field of building energy and building performance simulation, when a machine training model is applied, enough review of data correlation is insufficient, and a review of input variable selection is not conducted in the HVAC&R system.

Thus, in this chapter, to create an ANN model, quantitative judgment on input variables for individual sub-models (models for each flowrate and model for each equipment) is performed to select appropriate variables. However, in the case of the hyper-parameter optimization process, a common level was applied to the number of hidden layers, nodes (neurons), and activation functions, which have a great influence on the review of all sub-models.

5.1 Overview of Input Variable Selection Method

In the simulation or modeling process, input variables have a great influence on the calculation results. Snieder[17] is a method of selecting an existing input variable and is largely divided into a model-free method and a model-based method. The model-free method is a method of analyzing errors by grasping the correlation between the input variable (independent variable) and the output variable (dependent variable) through a statistical approach. In the model-based method, input variables are sequentially added (Forward) or sequentially removed from all input variables (Backward) and substituted into the model for calculation. After that, the model with the best performance is selected and the input variable at that time is adopted.

However, with the model-based method, the computational cost is high in the process of responding to all the variables, and there is a risk of overfitting. In addition, when the number of initial variables is large, there is a disadvantage in that the process of finding a suitable combination set is complicated. We selected an understanding model-free technique to make the sub-model efficiently and aimed to be systematically established this method.

The correlation coefficient between two variables is used as a statistical-based input variable selection method applied in this study. There are Pearson and Spearman methods for calculating the correlation coefficient, and the relationship between the two variables is grasped, and the degree is grasped by the coefficient at that time.

The Pearson correlation coefficient is a measure of the linear relationship between two variables. When one variable increase and the other variable increases, it has a positive correlation (+1), and when one variable increase and the other variable decreases, it has a negative correlation (-1). If there is no linear relationship between the two variables, it is marked as 0. Therefore, the Pearson correlation coefficient has a value from -1 to +1, and in this study, it was determined that the negative correlation also affects the creation of the ANN model.

In the case of Spearman, this is a method of grasping the relationship between two variables by grasping the non-linear relationship by a non-parametric method. The Spearman correlation coefficient determines the monotony of whether another variable simply increases or decreases when one variable increase and ranks the values to calculate the correlation coefficient for the ranking. Therefore, the correlation coefficient can be calculated even when the data is not a continuous variable. Only Spearman also has a negative value when the relationship decreases, and in this study, the absolute value was taken and the input variables were judged in the order of the highest degree of relationship.

The method is also important in selecting the input variable,

Table 1. Case setting

Pearson Coefficient		Spearman Coefficient	
Case	Criterion	**Case**	Criterion
Case01	80% of Rank	**Case06**	80% of Rank
Case02	60% of Rank	**Case07**	60% of Rank
Case03	40% of Rank	**Case08**	40% of Rank
Case04	20% of Rank	**Case09**	20% of Rank
Case05	10% of Rank	**Case10**	10% of Rank
Default	Total	Physics variables case	

Table 2. Total input variables by flowrate model

Count	Input Variables	Unit
01	Time and minute	-
02	Week	-
03	Outdoor Temperature	℃
04	Outdoor Relative Humidity	%
05	Cooling Load	kW
06	Heat Storage Capacity	kW
07-08	Outlet Chilled Water Temperature before 1minute (Refrigerator 01, 02)	℃
09-10	Inlet Chilled Water Temperature (Refrigerator 01, 02)	℃
11	Outlet Low-Temperature of HEX before 1minute	℃
12	Inlet Low-Temperature of HEX	℃
13	Outlet High-Temperature of HEX before 1minute	℃
14	Inlet High-Temperature of HEX	℃
15	Outlet Low-Temperature of Tank before 1minute	℃
16	Inlet Low-Temperature of Tank	℃
17	Outlet High-Temperature of Tank before 1minute	℃
18	Inlet High-Temperature of Tank	℃
19	Return Header Temperature before 1minute	℃
20	Supply Header Temperature	℃

but more importantly, it can be said that the criteria, that is, the accuracy of the machine training model varies depending on the level of the input variable to which the input variable is used. When using statistical techniques, a set of input variables must be configured according to certain criteria.

In the case of the Pearson and Spearman correlation coefficient, ±0.4~0.6 indicates a somewhat high correlation, ±0.6~0.8 indicates a high correlation, and ±0.8 or more indicates a very high correlation, and values below or above have low or no correlation. However, when selecting input variables, there is no clear criterion as to which level will be specifically selected as the criteria. In addition, there is a problem that there may be a case in which the values of the Pearson and Spearman correlation coefficients may not always have more than ±0.8.

The accuracy of the model was calculated by grouping variables up to the level corresponding to the top 10, 20, 40, 60, and 80% according to the ranking result of the correlation coefficient values of all variables. We put these as five cases, and a total of ten cases were generated for each of the Pearson and Spearman correlations (Table 1). In addition, the review of all input variables that are candidate groups and the case of the equipment model was reviewed by adding a knowledge-based input variable selection case that reflects the physical characteristics of the heat source system.

5.2 Overview of Input Variable Selection in Flow Prediction Model

Among the total collected data, candidate input variables were selected based on variables that directly affect the operation of the system. In the case of the flowrate prediction model, the purpose is to identify the flowrate required for water supply by each piece of equipment in the heat source system. In other words, it attempts to accurately identify the flowrate to be sent to the air conditioning system to handle the load. After selecting items that affect the water supply flowrate for each piece of equipment, detailed input variables were selected.

Items that affect the water supply flowrate for each piece of equipment include operation time, outdoor air condition, inlet temperature for each piece of equipment, performance for each equipment, and outlet temperature for each piece of equipment.

Table 3. Total input variables by equipment model

Count	Input Variables	Unit
01	Time and minute	-
02	Week	-
03	Outdoor Temperature	℃
04	Outdoor Relative Humidity	%
05	Cooling Load	kW
06	Heat Storage Capacity	kW
07-11	Flowrate of each Pump	m^3/h
12-13	Inlet Chilled Water Temperature (Refrigerator 01, 02)	℃
14	Inlet Low-Temperature of HEX	℃
15	Inlet High-Temperature of HEX	℃
16	Inlet Low-Temperature of Tank	℃
17	Inlet High-Temperature of Tank	℃
18	Supply Header Temperature	℃

Details are shown in Table 2, and input variables were selected and modeled for each of the five variable flow pumps for a total of 20 items. Concretely, 31 input variables for the flowrate of a total of five pumps were investigated. Table 5 shows the value of the Pearson and Spearman correlation coefficient and the ranking by item accordingly.

Calculating the Pearson correlation coefficient, the cooling

Table 5. Calculation of correlation coefficient for each input variable and ranking result by flowrate model

Coefficient	Input Variables		Week	Tank Storage Rate	Time and Minute	Outlet Low-Temperature of Tank (1minute ago)	Inlet Low-Temperature of Tank	Outlet High-Temperature of Tank (1minute ago)	Inlet High-Temperature of Tank	Supply Header Temperature	Cooling Load	Return Header Temperature (1minute ago)	Outlet Water Temperature of Chiller02 (1minute ago)	Inlet Chilled Water Temp. of Chiller02	Outlet Water Temperature of Chiller01 (1minute ago)	Inlet Chilled Water Temperature of Chiller01	Outdoor Temperature	Outdoor Relative Humidity	Outlet Low-Temperature of Tank (1minute ago)	Inlet Low-Temperature of HEX	Outlet High-Temperature of Tank (1minute ago)	Inlet High-Temperature of HEX
Pearson Coefficient	Flowrate 01	Value	0.05	0.62	0.01	0.12	0.11	0.01	0.19	0.16	0.91	0.15	0.15	0.42	0.71	0.42	0.38	0.30	0.47	0.41	0.43	0.09
		Rank	18	3	20	15	16	19	11	12	1	13	14	7	2	6	9	10	4	8	5	17
	Flowrate 02	Value	0.13	0.83	0.23	0.02	0.69	0.22	0.45	0.52	0.46	0.20	0.83	0.25	0.31	0.06	0.37	0.39	0.50	0.62	0.67	0.28
		Rank	18	1	15	20	3	16	9	6	8	17	2	14	12	19	11	10	7	5	4	13
	Flowrate 03	Value	0.01	0.88	0.10	0.25	0.30	0.15	0.29	0.28	0.95	0.16	0.49	0.45	0.88	0.50	0.44	0.46	0.78	0.58	0.71	0.06
		Rank	20	2	18	15	12	17	13	14	1	16	8	10	3	7	11	9	4	6	5	19
	Flowrate 04	Value	0.01	0.88	0.10	0.25	0.30	0.15	0.29	0.28	0.95	0.16	0.49	0.45	0.88	0.50	0.44	0.46	0.78	0.58	0.71	0.06
		Rank	20	2	18	15	12	17	13	14	1	16	8	10	3	7	11	9	4	6	5	19
	Flowrate 05	Value	0.02	0.84	0.07	0.21	0.25	0.09	0.27	0.26	1.00	0.16	0.39	0.47	0.86	0.50	0.45	0.43	0.72	0.56	0.66	0.06
		Rank	20	3	18	15	14	17	12	13	1	16	11	8	2	7	9	10	4	6	5	19
Spearman Coefficient	Flowrate 01	Value	0	0.7	0.1	0.2	0.3	0.2	0.3	0.2	0.8	0.1	0.4	0.4	0.7	0.6	0.4	0.4	0.6	0.5	0.4	0.1
		Rank	20	2	19	15	12	14	13	16	1	17	11	9	3	4	8	10	5	6	7	18
	Flowrate 02	Value	0.1	0.7	0.2	0	0.8	0.2	0.4	0.5	0.5	0.2	0.9	0.1	0.2	0.1	0.3	0.4	0.4	0.6	0.6	0.3
		Rank	18	3	13	20	2	14	8	7	6	15	1	19	16	17	11	10	9	5	4	12
	Flowrate 03	Value	0	0.9	0.1	0.2	0.5	0.3	0.4	0.4	1	0	0.6	0.4	0.6	0.5	0.5	0.5	0.8	0.7	0.8	0.1
		Rank	19	2	17	16	10	15	13	12	1	20	7	14	6	8	11	9	3	5	4	18
	Flowrate 04	Value	0	0.9	0.1	0.2	0.5	0.3	0.4	0.4	1	0	0.6	0.4	0.6	0.5	0.5	0.5	0.8	0.7	0.8	0.1
		Rank	19	2	17	16	10	15	13	12	1	20	7	14	6	8	11	9	3	5	4	18
	Flowrate 05	Value	0	0.9	0.1	0.2	0.5	0.3	0.4	0.5	1	0	0.5	0.4	0.6	0.5	0.5	0.5	0.8	0.7	0.8	0.1
		Rank	19	2	17	16	11	15	13	12	1	20	7	14	6	8	10	9	3	5	4	18

load and the heat storage amount of the heat storage tank for all five pump flowrates ranked at the top in correlation with each flowrate. This is a sufficiently reasonable ranking since the processing load (cooling load) and processing capacity (heat storage capacity), which have the greatest effect on the operation of the system, have a great influence on the operation of each piece of equipment. In addition, it was confirmed that the temperature of the chilled water supplied from the equipment connected to the pump is also directly related to the flowrate. The conditions for outside temperature, humidity, and time are in the lower zone, so it was judged that the influence on the operation of the heat source system was low. Even though the HVAC&R system is time-series data, information on time were placed in a subordinate priority category when selecting an input variable. For the five cases, the input variables occupying 80, 60, 40, 20, and 10% from the top rank were sequentially grouped and calculated according to the case setting method.

Calculating the Spearman correlation coefficient, the high rankings show similar patterns to the Pearson correlation coefficient results. Also, it was found that there was a high monotonic correlation between the cooling load and the amount of heat storage, and the outside temperature, humidity, and visual information were ranked second. However, the specific ranking clearly differs from the result of the Pearson correlation coefficient, and the discernment for this should be judged according to the calculation result for each case in Chapter 6.

Table 4. Physics variables of each equipment

Equipment	Input Variables of Physics Case	Unit
Air-Cooled Turbo Refrigerator	Inlet Water Temperature	°C
	Inlet Air Temperature	°C
	Flow Rate of Water	m³/h
Pump	Inlet Water Temperature	°C
	Flow Rate	m³/h
Heat Storage Tank	Flow Rate of Low-Temperature	m³/h
	Flow Rate of Low-Temperature	m³/h
	Inlet Low-Temperature	°C
	Inlet High-Temperature	°C

5.3 Overview of input variable selection of equipment operation model

In the case of the equipment model, a parallel calculation is performed during the simulation of the entire system. It is also important to review individual input variables for each equipment, but before that, a group of input variables that can encompass the entirety of the equipment is required. In the case of the equipment model, the output variable has an indicator for power consumption and performance, and the outlet temperature. Therefore, we expected the power consumption and the outlet temperature in the input variables.

Table 3 shows the common candidate group items for all

Table 6. Calculation of correlation coefficient for each input variable and ranking result by refrigerator model

Coefficient	Input Variables	Week	Tank Storage Rate	Time and Minute	Flowrate of Tank	Inlet Low-Temperature of Tank	Inlet Low-Temperature of Tank	Supply Header Temperature	Cooling Load	Flowrate of Chiller02	Inlet Water Temperature of Chiller02	Flowrate of Chiller01	Inlet Water Temperature of Chiller01	Outdoor Temperature	Outdoor Relative Humidity	Flowrate of HEX	Inlet Low-Temperature of HEX	Inlet High-Temperature of HEX	Flowrate of Secondary Pump
Pearson Coefficient	R01 Tpara	0.05	0.62	0.00	0.74	0.13	0.18	0.15	0.90	0.30	0.43	0.99	0.41	0.42	0.33	0.74	0.40	0.10	0.89
	Rank	17	6	18	4	15	13	14	2	12	7	1	9	8	11	5	10	16	3
	R02 Tpara	0.03	0.78	0.06	0.94	0.21	0.21	0.15	0.96	0.41	0.47	0.84	0.51	0.42	0.41	0.94	0.46	0.13	0.95
	Rank	18	6	17	4	13	14	15	1	12	8	5	7	10	11	3	9	16	2
	R01 COPpara	0.15	0.81	0.11	0.48	0.65	0.47	0.51	0.41	0.94	0.13	0.24	0.00	0.27	0.33	0.48	0.60	0.30	0.43
	Rank	15	2	17	6	3	8	5	10	1	16	14	18	13	11	6	4	12	9
	R02 COPpara	0.13	0.54	0.24	0.24	0.62	0.33	0.38	0.08	0.79	0.09	0.17	0.08	0.21	0.28	0.24	0.39	0.26	0.10
	Rank	14	3	11	9	2	6	5	18	1	16	13	17	12	7	9	4	8	15
Spearman Coefficient	R01 Tpara	0	0.75	0.06	0.83	0.31	0.25	0.19	0.84	0.29	0.41	1.00	0.63	0.43	0.41	0.83	0.45	0.07	0.84
	Rank	18	6	17	4	12	14	15	2	13	10	1	7	9	11	4	8	16	3
	R02 Tpara	0	0.75	0.06	0.83	0.31	0.25	0.19	0.84	0.30	0.41	1.00	0.64	0.42	0.40	0.83	0.45	0.07	0.84
	Rank	18	6	17	4	12	14	15	2	13	10	1	7	9	11	4	8	16	3
	R01 COPpara	0.1	0.76	0.17	0.52	0.79	0.45	0.50	0.49	0.97	0.01	0.28	0.10	0.26	0.32	0.52	0.56	0.32	0.49
	Rank	16	3	15	5	2	10	7	9	1	18	13	17	14	11	5	4	12	8
	R02 COPpara	0.1	0.60	0.28	0.40	0.75	0.39	0.44	0.36	0.93	0.02	0.14	0.05	0.23	0.30	0.40	0.46	0.31	0.36
	Rank	15	3	13	6	2	8	5	10	1	18	16	17	14	12	6	4	11	9

equipment. For a total of 18 candidate group input variables, cases were set according to the criteria in the same manner as the flowrate model. In addition, in the case of the equipment model, input variables based on the theory of the first law were additionally reviewed, and the input variables based on the theoretical equations for each equipment are shown in Table 4.

5.3.1 Calculation result of correlation coefficient of air-cooled turbo chiller

Parameter 01 for the chilled water outlet temperature and COP parameter 02 for calculating the power amount and performance were derived for a chiller. In the target heat source system, there are two chillers (chiller 01 and 02) and correlation analysis for all candidate input variables for each of the two parameters is carried out for each chiller. Table 6 shows the Pearson and Spearman correlation coefficient values and rankings.

Calculating the correlation coefficient between Pearson and Spearman, it was determined that the cooling load and the flowrate of the equipment had the highest influence in the case of the chilled water outlet temperature parameter (i.e. Tpara in Fig.6). In addition, the flowrate of chilled water returned to the equipment was also found to have a high effect. For the COP parameter (i.e. COPpara in Fig.6), the flowrate of the turbo chiller 02 for heat storage operation, the inlet temperature at the low-temperature side of the heat storage tank, and the amount of heat storage were equally high. Both types of parameters were found to be unaffected by driving time and outside air conditions.

Table 8. Calculation of correlation coefficient for each input variable and ranking result by Tank Model

Input Variables	Week	Tank Storage Rate	Time and Minute	Flowrate of Tank	Inlet Low-Temperature of Tank	Inlet Low-Temperature of Tank	Supply Header Temperature	Cooling Load	Flowrate of Chiller02	Inlet Water Temperature of Chiller02	Flowrate of Chiller01	Inlet Water Temperature of Chiller01	Outdoor Temperature	Outdoor Relative Humidity	Flowrate of HEX	Inlet Low-Temperature of HEX	Inlet High-Temperature of HEX	Flowrate of Secondary Pump
Pearson	0.29	0.01	0.07	0.02	0.19	0.20	0.24	0.00	0.06	0.22	0.04	0.21	0.05	0.11	0.02	0.17	0.23	0.01
Rank	1	16	10	14	7	6	2	18	11	4	13	5	12	9	14	8	3	17
Spearman	0.25	0.07	0.08	0.01	0.14	0.28	0.31	0.03	0.05	0.29	0.07	0.30	0.03	0.09	0.01	0.24	0.28	0.03
Rank	6	12	10	17	8	4	1	15	13	3	11	2	14	9	17	7	5	16

Table 9. Calculation of correlation coefficient for each input variable and ranking result by pump model

Coefficient	Input Variables	Week	Tank Storage Rate	Time and Minute	Flowrate of Tank	Inlet Low-Temperature of Tank	Inlet Low-Temperature of Tank	Supply Header Temperature	Cooling Load	Flowrate of Chiller02	Inlet Water Temperature of Chiller02	Flowrate of Chiller01	Inlet Water Temperature of Chiller01	Outdoor Temperature	Outdoor Relative Humidity	Flowrate of HEX	Inlet Low-Temperature of HEX	Inlet High-Temperature of HEX	Flowrate of Secondary Pump
Pearson Coefficient	Pump 01	0.01	0.63	0.09	0.76	0.21	0.15	0.08	0.61	0.35	0.30	0.32	0.41	0.23	0.33	0.76	0.34	0.11	0.61
	Rank	18	3	16	2	13	14	17	5	7	11	10	6	12	9	1	8	15	4
	Pump 02	0.02	0.07	0.15	0.12	0.31	0.09	0.11	0.33	0.33	0.10	0.58	0.19	0.02	0.06	0.12	0.02	0.15	0.31
	Rank	18	14	7	10	5	13	11	3	2	12	1	6	16	15	9	17	8	4
	Pump 03	0.02	0.50	0.19	0.35	0.41	0.36	0.53	0.27	0.52	0.15	0.13	0.03	0.35	0.39	0.35	0.53	0.29	0.31
	Rank	18	4	14	9	5	7	1	13	3	15	16	17	10	6	8	2	12	11
	Pump 04	0.02	0.55	0.19	0.41	0.43	0.38	0.56	0.32	0.55	0.17	0.17	0.05	0.37	0.41	0.41	0.57	0.29	0.35
	Rank	18	3	14	8	5	9	2	12	4	15	16	17	10	6	7	1	13	11
	Pump 05	0.02	0.47	0.18	0.31	0.40	0.36	0.53	0.24	0.50	0.13	0.10	0.01	0.33	0.37	0.31	0.52	0.30	0.27
	Rank	17	4	14	9	5	7	1	13	3	15	16	18	8	6	10	2	11	12
Spearman Coefficient	Pump 01	0	0.7	0.1	0.8	0.3	0.2	0.2	0.8	0.4	0.4	0.9	0.6	0.4	0.4	0.8	0.4	0.1	0.8
	Rank	18	6	16	2	13	14	15	4	12	10	1	7	11	9	2	8	17	5
	Pump 02	0.1	0.5	0.3	0.4	0.7	0.4	0.5	0.3	0.9	0	0.1	0	0.2	0.3	0.4	0.4	0.4	0.3
	Rank	15	3	13	8	2	6	4	11	1	17	16	18	14	12	8	5	7	10
	Pump 03	0	0.7	0.2	0.8	0.6	0.4	0.5	0.8	0.6	0.2	0.4	0.2	0.4	0.5	0.8	0.7	0.2	0.8
	Rank	18	5	17	1	8	12	9	4	7	16	13	14	11	10	1	6	15	3
	Pump 04	0	0.7	0.2	0.8	0.6	0.4	0.6	0.8	0.6	0.2	0.4	0.2	0.4	0.5	0.8	0.7	0.2	0.8
	Rank	18	5	17	1	9	12	8	4	7	16	13	14	11	10	1	6	15	3
	Pump 05	0	0.7	0.2	0.8	0.5	0.4	0.6	0.8	0.6	0.2	0.4	0.2	0.4	0.5	0.8	0.7	0.3	0.8
	Rank	18	5	17	2	9	11	7	4	8	16	13	15	12	10	2	6	14	1

In the case of the chiller model, since there are input variables that can be inferred from the theoretical equations, a review was also conducted. There are three input variables based on the physical formula of the air-cooled turbo chiller: the equipment inlet chilled water temperature, the outside air temperature (heat source side inlet temperature), and the chilled water flowrate. We set these three input variables as a physics case.

5.3.2 Calculation result of correlation coefficient of Heat Storage Tank

Calculating the correlation coefficient for the parameters of the heat storage tank (Table 8), all the input variable candidate groups had relatively low coefficient values compared to other sub-models. In the case of other sub-models, there was an input variable item with a correlation coefficient value of ±0.8 or more, but in the case of the heat storage tank model, there was no input variable whose maximum value exceeded ±0.4 in both the Pearson and Spearman methods. As input variables based on the physical equation, inlet temperature and flowrate of the high and low-temperature sides of the heat storage tank, a total of four, were set as one set.

5.3.3 Calculation result of correlation coefficient of Variable Flowrate Pump

Calculating the correlation for each input variable for each of the five parameters of variable flow pumps (Table 9), it was found that all the input variables ranked in the top ranks had individual characteristics.

Spearman correlation coefficient has a high correlation between the flowrate and temperature of the equipment connected to the pump, indicating that sufficiently predictable variables are sequentially selected. However, in the case of the Pearson correlation coefficient, the flowrate and temperature of the indirectly connected equipment have higher coefficients than the information about the directly connected equipment. This can cause reasonable doubt as to whether it is an appropriate method to grasp an input variable with a simple linear relationship, and there is a need to determine the degree of influence of each input variable through the modeling results in Chapter 6.

RMSE	0.92	5.78	2.02	1.79	1.73	1.19	5.71	1.98	1.89	1.73	1.12	5.42	1.93	1.73	1.74	1.31	5.34	2.15	1.95	1.80	0.75	8.90	2.32	1.95	1.74
MAE	0.27	2.07	0.82	0.67	0.99	0.35	1.80	0.72	0.70	0.98	0.32	1.77	0.75	0.68	0.98	0.33	1.76	0.80	0.72	1.03	0.23	4.09	1.04	0.75	0.98
MBE	0.03	1.60	0.17	0.09	0.09	0.25	-0.17	-0.04	-0.03	0.04	-0.03	0.43	0.20	0.27	-0.05	0.13	0.63	0.05	0.09	0.36	0.08	0.44	0.74	0.32	0.05

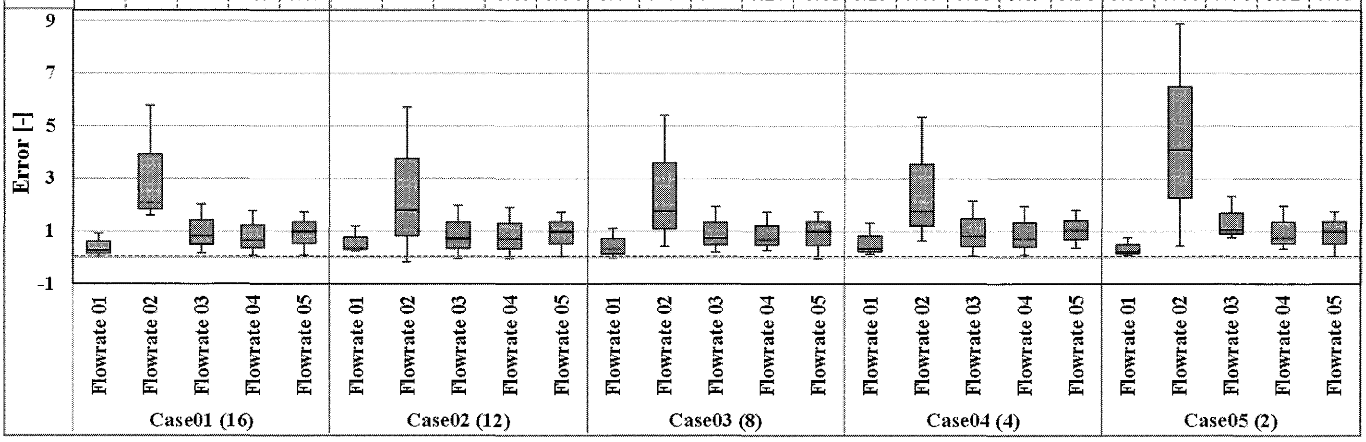

Fig. 3 Results of flowrate model using Pearson correlation coefficient

RMSE	1.07	5.09	1.90	1.74	1.73	1.10	6.10	1.96	1.83	1.73	1.22	5.96	2.09	2.17	1.77	1.44	5.13	2.19	2.01	1.78	1.52	2.83	2.32	2.02	1.90
MAE	0.32	1.63	0.69	0.62	0.98	0.32	1.93	0.71	0.68	0.98	0.35	1.88	0.96	1.07	1.01	0.38	1.61	0.89	0.82	1.01	0.44	0.57	1.05	0.83	1.07
MBE	0.16	1.15	0.02	0.00	-0.02	0.18	0.51	-0.10	-0.02	-0.06	0.26	0.76	0.71	0.90	0.34	0.24	0.68	0.52	0.47	0.10	0.17	0.46	0.75	0.47	0.43

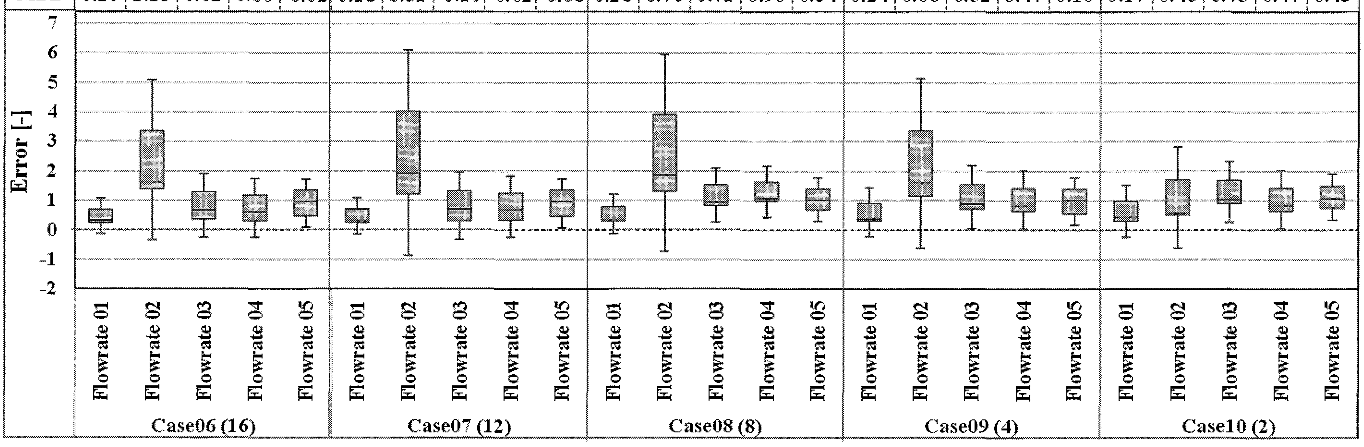

Fig. 4 Results of flowrate model using Spearman correlation coefficient

6. Validation of model accuracy

As means to measure the accuracy of the model, the following three were used in this study. Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Mean Bias Error (MBE). The results directly calculated from the ANN model, RMSE, MAE, and MBE were used to determine the case with the least error. Also, the number in parentheses of each case means the number of meanings entered for each case.

6.1 Flowrate Model

First, the calculation results of the ANN model of the flowrate model are shown in Fig. 3 and 4 for the Pearson correlation coefficient (cases 01-05) and Spearman correlation coefficients (cases 06-10). For all cases, five cases in which input variables compared to the Pearson correlation coefficient had high accuracy. Grasping the input variable as a linear relationship for the flowrate value, it is more advantageous to identify the variable with a higher correlation than to grasp the input Spearman correlation coefficient variable as a non-linear.

For the flowrate of five pumps, pump 02 is a pump of a chiller for heat storage operation, and the accuracy of pump 02 was relatively low in all ten cases. The flowrate of pump 02 transmits and receives a constant flowrate during night heat storage regardless of time and load, and it is considered that the classification model calculation method is more suitable than the regression model. That is, the flowrate during heat storage operation and the flowrate during other operations should be classified and calculated by the classification model. However, the ANN model applied in this study is a model suitable for calculating the regression equation, and it was determined that the accuracy of the flowrate of pump 02 having a constant value was lower than that of other pumps with fluctuating flowrates.

When selecting the input variable according to the value of the Pearson correlation coefficient, cases 01-04 had similar accuracy for five flowrates. Cases 03 and 04 generally had high accuracy for the flowrate of five pumps. This suggests that when an input variable with low correlation is added, the

RMSE	0.43	0.68	1.48	0.99	0.37	0.68	1.12	0.89	0.38	0.72	1.56	0.81	0.41	0.73	1.66	0.83	0.36	0.94	1.45	0.85	0.39	0.56	1.53	0.88
MAE	0.14	0.32	0.54	0.43	0.11	0.31	0.40	0.28	0.12	0.32	0`.62	0.23	0.12	0.32	0.64	0.21	0.09	0.44	0.52	0.28	0.10	0.23	0.56	0.31
MBE	0.13	0.27	-0.54	0.31	0.04	0.13	-0.18	0.02	0.05	0.15	-0.58	-0.04	0.06	0.14	-0.63	-0.09	0.06	0.38	-0.52	0.23	0.09	0.20	-0.56	0.27

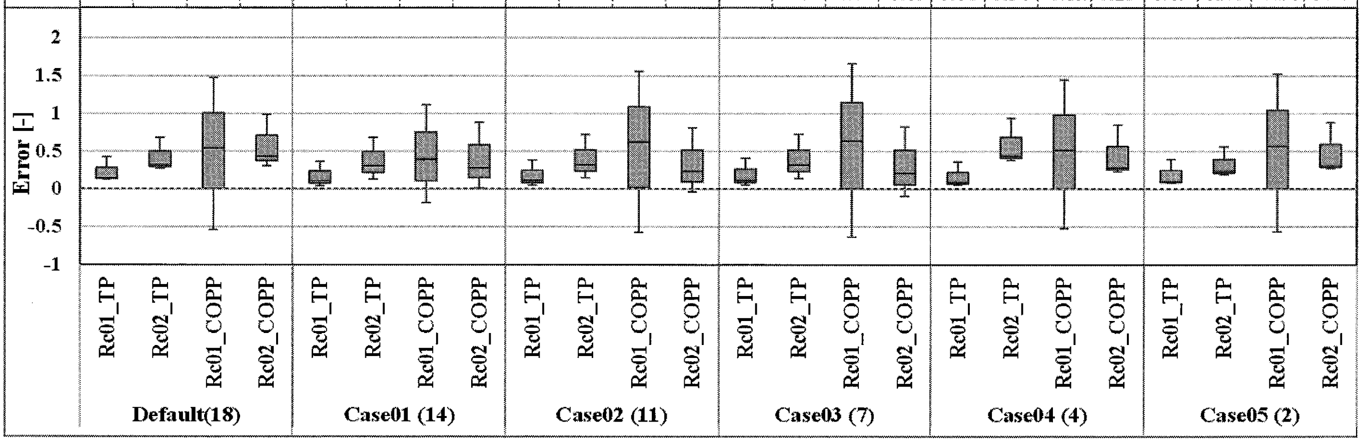

Fig. 5 Results of refrigerator parameter using Pearson correlation coefficient

RMSE	0.40	0.51	1.76	1.27	0.37	1.41	1.65	0.89	0.39	0.77	1.57	0.87	0.40	0.78	1.72	0.72	0.34	0.70	1.57	0.69	0.32	0.66	1.55	1.05
MAE	0.09	0.26	0.68	0.65	0.11	0.68	0.65	0.28	0.12	0.35	0.60	0.24	0.12	0.35	0.64	0.26	0.09	0.32	0.58	0.20	0.08	0.30	0.59	0.51
MBE	0.07	0.25	-0.68	0.53	0.04	-0.05	-0.64	0.01	0.07	0.23	-0.59	-0.05	0.04	0.23	-0.64	0.10	0.07	0.17	-0.58	-0.02	0.04	0.29	-0.59	0.41

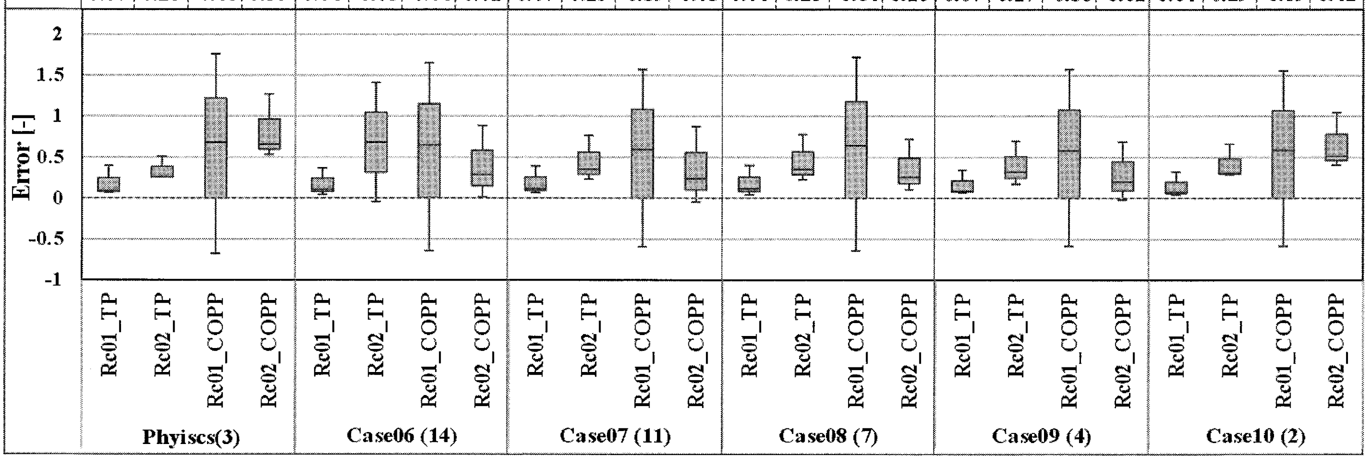

Fig. 6 Results of refrigerator parameter using Spearman correlation coefficient

accuracy of the calculation may be lowered. In the case of the Spearman correlation coefficient case, the error of case 10 with two input variables was the lowest.

In the case of the flowrate model, it is judged that input variables of a certain level (about 50% or more in this study) are necessary because various input variables have a complex effect on the flowrate. It is judged that it is necessary to improve the performance of the entire model by adding a process for optimizing hyper-parameters of the ANN model in the future.

6.2 Equipment Model

6.2.1 Air-cooled Turbo Chiller Model

The results of calculating the two parameters (chilled water

RMSE	99.63	64.37	78.16	80.93	75.68	62.11	99.49	68.89	89.90	69.22	86.06	347.35
MAE	57.05	37.67	40.11	39.40	39.89	30.57	50.82	34.35	46.15	37.02	47.51	212.03
MBE	50.18	28.83	26.30	19.39	22.29	18.99	38.09	19.86	20.25	21.27	26.03	-10.35

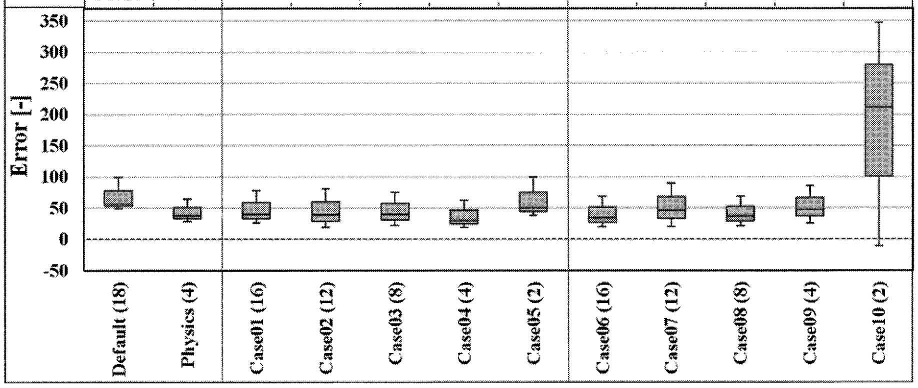

Fig. 7 Results of heat storage tank parameter by each case

RMSE	1.22	1.19	2.82	2.00	3.57	7.78	0.76	9.70	7.86	9.88	4.32	0.66	4.51	3.06	4.54	4.22	0.86	3.99	2.09	4.58	1.34	0.63	2.77	1.62	3.21	1.18	1.13	3.10	1.87	3.19
MAE	0.33	0.51	1.29	1.08	1.47	2.88	0.27	3.70	3.33	3.79	1.17	0.15	1.64	1.36	1.62	0.99	0.23	1.42	1.04	1.56	0.36	0.20	0.89	0.82	1.11	0.31	0.41	1.24	0.98	1.26
MBE	0.03	0.47	1.20	0.82	1.27	1.61	0.25	2.72	2.49	2.63	0.35	0.12	1.19	0.79	1.14	0.24	0.22	0.94	0.65	1.25	0.06	0.18	0.67	0.54	0.88	-0.06	0.37	0.95	0.78	0.93

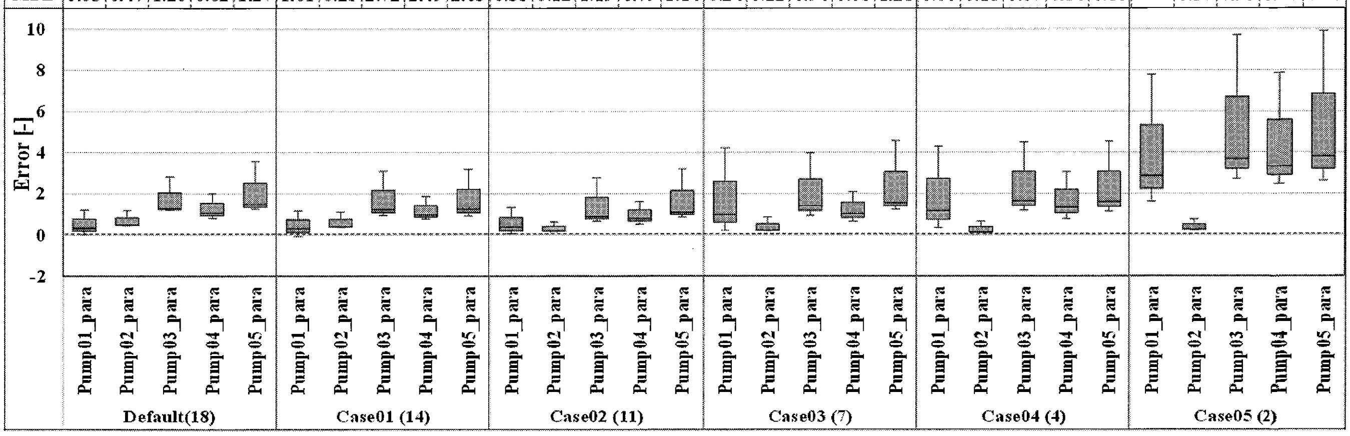

Fig. 8 Results of pump parameter using Pearson correlation coefficient

RMSE	1.08	0.58	3.48	2.23	4.39	1.58	0.46	12.39	10.13	3.47	0.98	0.96	1.44	0.72	1.47	1.11	1.05	2.67	1.74	2.25	1.13	1.02	2.62	1.31	2.53	1.23	1.55	2.33	1.58	2.70
MAE	0.34	0.29	0.92	0.80	1.39	0.43	0.22	6.34	5.56	1.12	0.32	0.43	0.56	0.41	0.66	0.26	0.31	1.12	0.90	0.86	0.32	0.30	1.02	0.58	1.01	0.35	0.54	0.88	0.75	0.90
MBE	-0.05	0.27	0.69	0.37	1.00	-0.01	0.21	1.91	1.33	0.82	0.02	0.34	0.37	0.33	0.42	0.05	0.21	0.89	0.70	0.60	-0.07	0.25	0.72	0.31	0.82	0.03	0.46	0.66	0.56	0.61

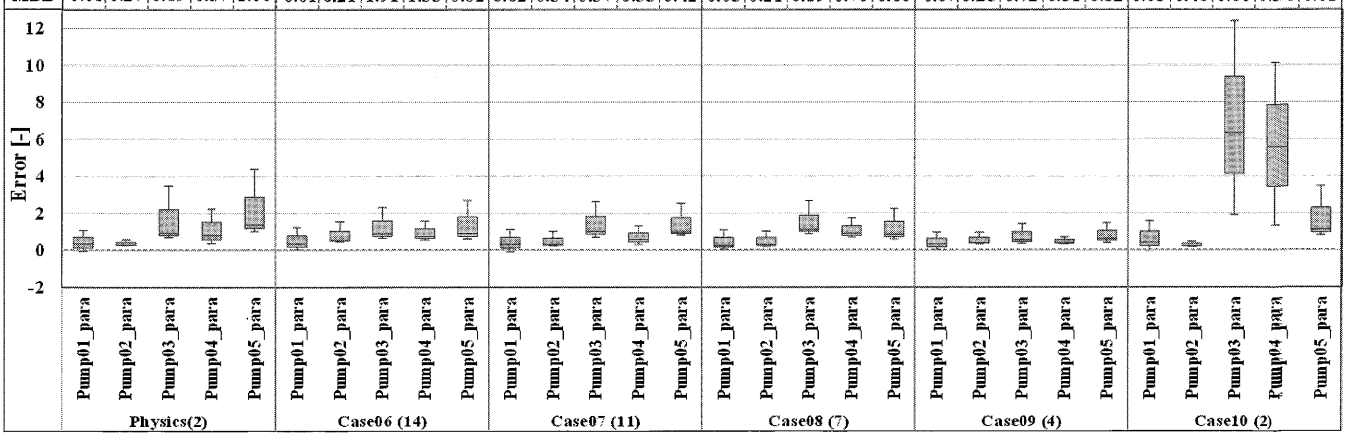

Fig. 9 Results of variable flowrate pump parameter using Spearman correlation coefficient

parameter, COP parameter) of the refrigerator using the ANN model are shown in Fig. 5 and 6. In the case of the chilled water parameter, the accuracy of the chiller for heat storage operation was high while the accuracy of the refrigerator for heat storage operation was high in the case of the COP parameter, but the accuracy of the refrigerator 01 handling the remaining load after the operation of the heat storage tank was low. This is a problem caused by the difference between the learning period and the prediction period in the operation pattern of the refrigerator 01.

6.2.2 Heat Storage Tank Model

Calculating the parameters of the heat storage tank model (Fig. 7), the error decreases as the number of input variables increases in both Pearson and Spearman correlation coefficients. It can be seen from Table 8 that the heat loss parameter of the heat storage tank has a low correlation for all input variables. Accordingly, when the number of input variables increased, the ANN model was calculated through a complex relationship between the input variables and the error of the model decreased. Also, like other equipment models, the case of input variables based on theoretical knowledge was found to have high accuracy.

6.2.3 Variable Flowrate Pump Model

The results of calculating the correlation of input variables for five variable flow pumps are shown in Fig. 8 and 9. When selecting input variables based on the Pearson correlation coefficient, the error decreased as the number of input variables increased. A simple linear relationship reduces the accuracy of the equipment model.

In the case of the Spearman correlation coefficient, except for case 10, the accuracy was higher than that of the Pearson correlation coefficient. In the case of case 10, even though the monotonic correlation was high, it was found that it was difficult to predict the correct parameter with only two input variables. However, in Cases 06 to 09, the difference in accuracy due to the number of input variables is insignificant.

When the detailed items in the heat source system are calculated using the ANN model, Spearman correlation coefficient has a higher overall effect than the Pearson correlation coefficient in determining the quantitative relationship of the input variable. This is because the operation of the heat source system operates by a non-linear relationship, so it can be said that the accuracy of the heat source system is simply degraded by a linear correlation. In addition, it has higher accuracy than when using only a small amount of input variables (2 to 3) when substituting input variables of about 50% of the upper rank.

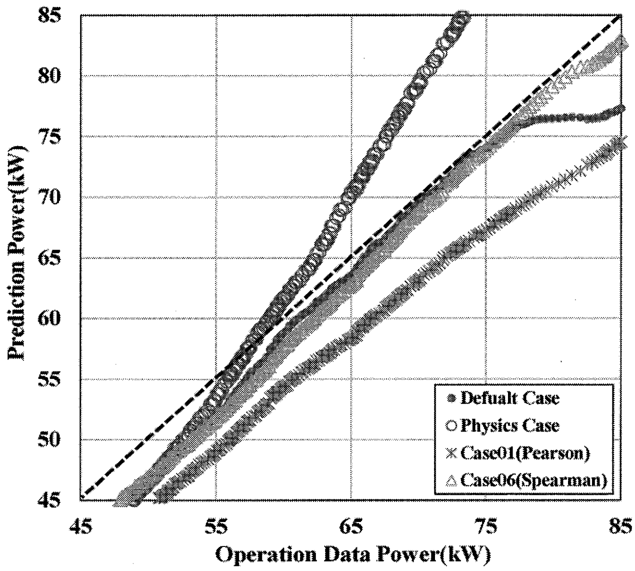

Fig. 10 Chiller 02 power consumption result according to 4 input variable cases

Table 10. The results of temperatures of the air-cooled refrigerator and heat storage tank

CV(RMSE) [%]		Chiller01 Outlet Temp.	Chiller02 Outlet Temp.	Layer01 Temp.	Layer10 Temp.
Default		4.47	4.19	0.72	1.52
Physics		3.17	3.56	0.72	1.52
Pearson Coefficient	Case01	3.75	4.63	0.72	1.52
	Case02	3.62	6.23	0.72	1.52
	Case03	4.60	5.46	0.72	1.51
	Case04	4.50	5.46	0.71	1.49
	Case05	4.45	5.27	0.71	1.50
Spearman Coefficient	Case06	2.99	5.01	0.72	1.52
	Case07	3.59	5.31	0.72	1.52
	Case08	4.43	5.49	0.72	1.51
	Case09	4.67	5.39	0.72	1.52
	Case10	4.45	9.87	0.73	1.54

Table 11. The result of calculating the power consumption of pumps

CV(RMSE) [%]		Pump01 Power	Pump02 Power	Pump03 Power	Pump04 Power	Pump05 Power
Default		21.56	25.43	36.99	35.65	34.96
Physics		16.44	16.20	14.04	15.11	23.70
Pearson Coefficient	Case01	17.17	22.05	30.15	32.05	32.25
	Case02	21.88	9.33	16.57	23.49	24.29
	Case03	36.91	9.71	33.44	28.29	36.18
	Case04	38.14	6.70	37.80	33.99	35.36
	Case05	122.55	12.68	55.74	59.25	59.27
Spearman Coefficient	Case06	20.08	35.92	19.52	20.23	17.89
	Case07	18.43	16.24	23.27	13.91	22.12
	Case08	15.99	27.19	29.76	28.04	20.10
	Case09	18.93	27.05	17.57	16.42	24.67
	Case10	19.25	11.02	152.22	147.84	23.47

6.3 Results of Calculating Equipment Performance

So far, the flowrate and parameters of the equipment were calculated using the ANN model. In the case of the equipment, it was required to examine the accuracy of the equipment performance according to the parameter by substituting the calculated parameter into the theoretical equation. In this section, the performance of the equipment is calculated using Coefficient of Variation of the Root Mean Square Error (CV(RMSE)). According to The American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE) guidelines[18], if the simulation results, CV(RMSE) satisfies ±30% (based on hourly data), it can be said to have enough accuracy. The calculation results for the outlet temperature and power amount for each equipment were reviewed as CV(RMSE) indicators.

Table 10 shows the CV(RMSE) for the results of the outlet temperature of the air-cooled refrigerator and the high and low temperatures of the heat storage tank. Both the results of the outlet temperature of the refrigerator and heat storage tank had a CV(RMSE) value that satisfies ±30%.

As the result of the power consumption of the refrigerator, as in the parameter result, refrigerator 01 had low accuracy and refrigerator 02 had high accuracy. In the case of Refrigerator 01, it was found that the CV(RMSE) of all cases was 200% or more, but in the case of Refrigerator 02, the CV (RMSE) of Case 06 using Spearman correlation coefficient was 16.3%, showing the highest performance. Fig. 10 shows the case in which all input variables are entered for the result of power consumption of chiller 02, (CV(RMSE): 20.4%), the case of variables based on physical equations (CV(RMSE): 24.1%), and Pearson coefficient case The results were shown in Case 01 with high performance (CV(RMSE): 22.6%) and Case 06 with the highest performance among Spearman correlation coefficient cases (CV(RMSE): 16.3%). As can be seen from the scatter plot, the higher the number of input variables, the higher the accuracy, and among them, the highest accuracy was obtained by sequentially entering the ranked input variables using the Spearman correlation coefficient.

Table 11 is the result of calculating the power consumption using the parameter calculation values of five pumps. Cases using physical equation-based variables and Case 02 applying Pearson correlation coefficient, and Cases 07 and 09 applying Spearman correlation coefficient satisfied CV(RMSE) 15% for all five pumps. It is judged that the accuracy of the parameters calculated using the ANN has a high influence on the power consumption of the pump. Therefore, it can be said that it has high accuracy to select and model comprehensive input variables after verifying whether the characteristics of physical variables are sufficiently reflected when calculating the input variables quantitatively.

7. Conclusion

This study proposed a new simulation model that can be modeled only with the data collected when performing a simulation on the heat source system of a non-residential building. In the case of the flowrate model, a black box method that directly calculates the flowrate value using the ANN model was applied, and the equipment model was developed for a new gray box method that calculated the physical equations after calculating the theoretical parameter by the ANN model.

In the development of the modeling method, we tried to create a high-precision model by judging the influence of the input variable. We were able to create a high-accuracy ANN model that sequentially selects input variables by applying Spearman correlation coefficients for both flow and equipment models. In the case of a heat source system, it is judged to be good to apply the Spearman coefficient, which is judged as a nonlinear monotonic composition, to correlate the input variables. In the case of the equipment model, it was highly accurate to examine the parameters based on the physical equations and the parameters to which the Spearman coefficient was applied, and then model them.

We will develop a hyper-parameter optimization to create an ANN model with higher accuracy and a connection model that can calculate the connection relationship between the model and the equipment, thereby completing the automatic simulation modeling method in which the simulation of the entire heat source system can be performed with data input only.

Reference

1) United Nations Environment Programme, 2019 Global Status Report for Buildings and Construction, pp.12-16, 2019.
2) ABB Group. Daikin Industries Ltd., Elster Group GmbH, GridPoint, Inc., Jones Lang LaSalle, Siemens AG (edited by). Energy Management Systems Market Size, Share & Trends Analysis Report by Product, By Component, By Solution, By Vertical, By End Use, By Region, And Segment Predicts, 2019 - 2025. Routledge. London (UK), 2019
3) F. Pugliese, F. De Paola, N. Fontana, M. Giugni, G. Marini, Experimental characterization of two Pumps As Turbines for hydropower generation. Renewable Energy, 99, pp.180–187, 2016
4) Y. Takagi, The Characteristic Analysis of a University Building: Dynamics Modelling by Grey Box Approach. SICE Journal of Control, 8, pp.328–333, 2015

5) J. E. Braun, N. Chaturvedi, An Inverse Grey-box Model for Transient Building Load Prediction. HVAC&R RESEARCH, 8, pp.73–99, 2002

6) T. Lee, W. Lu, An evaluation of empirically-based models for predicting energy performance of vapor-compression water chillers. Applied Energy, 87, pp.3486-3493, 2010

7) S. Estrada-Flores, I. Merts, B. De Ketelaere, J. Lammertyn, Development and validation of "grey-box" models for refrigeration applications: a review of key concepts. International Journal of Refrigeration, 29, pp.931-946, 2006

8) M. Hydeman, N. Webb, P. Sreedharan, S. Blanc, Development and Testing of a Reformulated Regression-Based Electric Chiller Model. 2002 Annual Conference, pp.1-10, 2002

9) M. R. Arahal, C. M. Cirre, M. Berenguel, Serial grey-box model of a stratified thermal tank for hierarchical control of a solar plant. Solar Energy, 82, pp.441-451, 2008

10) F. Berkenkamp, M. Gwerder, Hybrid model predictive control of stratified thermal storages in buildings. Energy and Buildings, 84, pp.233-240, 2014

11) G. Tardioli, A. Narayan, R. Kerrigan, M. Oates, J. O'Donnell, D. P. Finn, A methodology for calibration of building energy models at district scale using clustering and surrogate techniques. Energy and Buildings, 226, pp.1-24, 2020

12) K. Singh, R. Das, A feedback model to predict parameters for controlling the performance of a mechanical draft cooling tower. Applied Thermal Engineering, 105, pp.519–530, 2016

13) D. P. Kingma, J. L. Ba, Adam: A Method for Stochastic Optimization, Computing Research Repository(CoRR), vol. abs/1412.6980, 2014

14) A. F. Agarap, Deep Learning using Rectified Linear Units(ReLU), CoRR, vol. arXiv:1803.08375, 2019

15) J. M. Gordon, K. C. Ng, Predictive and diagnostic aspects of a universal thermodynamic model for chillers. International Journal of Heat and Mass Transfer, 38, pp.807-818, 1995

16) TRNSYS 18 Volume 4 Mathematical Reference, The Board of Regents of the University of Wisconsin System, 2017

17) E. Snieder, R. Shakir, U.T.Khan, A comprehensive comparison of four input variable selection methods for artificial neural network flow forecasting models. Journal of Hydrology, 583, pp.1-17, 2020

18) ASHRAE, ASHRAE Guideline 14-2002: Measurement of Energy Demand and Savings. Atlanta: American Society of Heating, Refrigerating and Air-Conditioning Engineers, 2002

Nomenclature

T_{cond}^{in} = Condenser Inlet Water Temperature (°C)
T_{evap}^{in} = Evaporator Inlet Water Temperature (°C)
Q_{evap} = Cooling load (kW)
$Q_{leak,eqv}$ = Evaporator Heat Leak (kW)
ΔS_T = Internal entropy generation (kW/K)
R = Thermal resistance (K/kW)
$T_{cond,TRNCata}^{in}$ = Condenser Inlet Water Temperature of TRNSYS Catalog (°C)
$Q_{evap,TRNCata}$ = Cooling load of TRNSYS Catalog (kW)
$COP_{TRNCata}$ = Coefficient of Performance from TRNSYS Catalog (-)
COP_{TRNSYS} = Coefficient of Performance from TRNSYS Simulation Result (-)
P_R = Power of Air-Cooled Chiller (kW)
$\dot{m}_{evap}$ = Flowrate into the Evaporator (m^3/h)
T_{evap}^{out} = Evaporator Outlet Water Temperature (°C)
COP = Calculated Coefficient of Performance (-)
T_a = Outdoor Temperature (°C)
$Power_{pump}$ = Power of Pump (kW)
η = Efficiency of Pump (%)
Q = Discharge Head (m^3/s)
ρ = Density of Water (kg/m^3)
H = Total Pump Head (m)
T_{pump}^{in} = Pump Inlet Water Temperature (°C)
$\dot{m}_{pump}$ = Flowrate into the Pump (m^3/h)
$\dot{m}_s$ = High-Temperature Side Flowrate (m^3/h)
T_{all} = Total Temperature of Heat Storage Tank (°C)
C_w = The specific heat of water (J/kg・K)
m = Total Flowrate of Heat Storage Tank (m^3/h)
$\dot{m}_e$ = Flowrate of each layer of Heat Storage Tank (m^3/h)
T_i = Temperature of each layer of Heat Storage Tank (°C)
$\frac{A_q \lambda_w}{z}$ = Heat Transfer Coefficient between layer (W/㎡·K)
$Tank_{SIT}$ = Inlet High-Temperature Side (°C)
U = Surface Heat Transfer Rate (W/㎡·K)
A_s = Surface Area (㎡)
n = The Number of Layer of Heat Storage Tank (-)

（受理：令和３年５月27日）